Ortholog prediction of the Aspergillus genus applicable for synthetic biology

Rasmussen, Jane Lind Nybo; Vesth, Tammi Camilla; Theobald, Sebastian; Frisvad, Jens Christian; Grigoriev, Igor V.; Baker, Scott E.; Andersen, Mikael Rørdam

Publication date:
2016

Document Version
Early version, also known as pre-print

Ortholog prediction of the Aspergillus genus applicable for synthetic biology

Jane Nybo (1), Tammi C. Vesth (1), Sebastian Theobald (1), Jens C. Frisvad (1), Igor V. Grigoriev (3), Scott E. Baker (2), Mikael R. Andersen (1)

1) Department of Systems Biology, Technical University of Denmark, Kgs. Lyngby, Denmark
2) Joint Bioenergy Institute, Berkeley, CA, USA
3) Joint Genome Institute, Walnut Creek, CA, USA

The Aspergillus genus contains leading industrial microorganisms, excelling in producing bioactive compounds and enzymes. Using synthetic biology and bioinformatics, we aim to re-engineer these organisms for applications within human health, pharmaceuticals, environmental engineering, and food production. In this project, we compare the genomes of +300 species from the Aspergillus genus to generate a high-resolution pan-genomic map, representing genetic diversity spanning ~200 million years. We are identifying genes specific to species and clades to allow for guilt-by-association-based mapping of genotype-to-phenotype. To achieve this, we have developed orthologous protein prediction software that utilizes genus-wide genetic diversity. The approach is optimized for large datasets, based on BLASTp considering protein identity and alignment coverage, and clustering using single linkage of bi-directional hits. The result is orthologous protein families describing the genomic and functional features of individual species, clades and the core/pan genome of Aspergillus; and applicable to genotype-to-phenotype analyses in other microbial genera.